

AN ANNOTATED BIBLIOGRAPHY OF  
EXPERT JUDGMENT IN ASSURANCE  
CASES

Patrick McGee  
Department of Philosophy, University of Florida

John Knight  
Department of Computer Science, University of Virginia

Department of Computer Science  
Technical Report CS-2016-02

5/31/2016

# An Annotated Bibliography of Expert Judgment in Assurance Cases

Patrick McGee

Department of Philosophy  
University of Florida  
Gainesville, FL  
p7rk@ufl.edu

John Knight

Department of Computer Science  
University of Virginia  
Charlottesville, VA  
knight@cs.virginia.edu

**Abstract.** Expert judgment is an important element of decision making in many areas of engineering. Although quantitative methods are usually preferred in assurance, necessary probabilities frequently cannot be determined and then engineers have to resort to the judgment of experts. Obviously, this situation leads to concern about how accurate the judgments of experts are. Accurate judgments are particularly important in systems for which the consequences of failure are high, such as security- and safety-critical systems. This report is an annotated bibliography on expert judgment.

1. Arnold, V., Clark, N., Collier, P. A., Leech, S. A., & Sutton, S. G. (2006). The Differential Use and Effect of Knowledge-Based System Explanations in Novice and Expert Judgment Decisions. *MIS Quarterly*, 30(1), 79–97. <http://www.jstor.org/stable/25148718>

Keyword(s): article, empirical studies, formation

While this article primarily focuses on knowledge-based systems, it does discuss a few ways in which expert thought processes differ from those of non-experts. “Choice patterns in using explanations indicated that novices used feedforward explanations more than experts did, while experts were more likely than novices to use feedback explanations. Novices also used more declarative knowledge and initial problem solving type explanations while experts used more procedural knowledge explanations. Finally, use of feedback explanations led to greater adherence to the KBS recommendations by experts--a condition that was even more prevalent as the use of feedback explanations increased” (79-80).

2. Bamber, E. M. (1983). Expert Judgment in the Audit Team: A Source Reliability Approach. *Journal of Accounting Research*, 21(2), 396–412. <http://www.jstor.org/stable/249078>

Keyword(s): article, empirical studies, modeling, probability

This article is about a study in which the sensitivity of audit managers' judgments to the reliability of the audit senior was examined. In the context of this study, the audit senior is viewed as a sort of expert who is relied upon by the audit manager. “Reliance on the audit senior occurs primarily through the recommendation of the senior that is expressed either explicitly... or implicitly... The manager may accept the recommendation of the senior, or s/he may

supplement that recommendation with a review of the relevant evidence...” (397). The article develops a probabilistic model of source reliability based on Bayes' Theorem and proceeds to describe the setup and results of an experiment designed to test the sensitivity of audit managers' judgments to the reliability of the audit senior. Based on the results, it is concluded that variations in source reliability have a significant effect on information content. Bamber believes that this “provides further evidence of the importance of context in auditor judgment and highlights the judgment dependencies existing in the audit team” (411).

3. Bengson, J. (2013). Experimental Attacks on Intuitions and Answers. *Philosophy and Phenomenological Research*, 86(3), 495–532. <http://doi.org/10.1111/j.1933-1592.2012.00578.x>

Keyword(s): article, empirical studies, formation

This article claims that a few commonly referenced studies for claims of the statistical irrelevance of expert opinion as an indicator of accuracy cannot definitely single out intuitions as source of research subjects' responses, and therefore cannot make any well-grounded conclusions about such intuitions. Bengson eventually goes on to argue for the claim that either the subjects' responses were not the result of intuitions or that incorrect responses actually conflicted with intuitions that were in fact, correct. Bengson recognizes that this claim isn't conclusive and reaches what he calls a “conciliatory position” which recognizes the utility of Experimental Philosophy, but advocates that such experimentation need be critically examined in order to determine what conclusions can be drawn from the data.

The final section of this article mentions several insightful considerations for future experiments of this kind in order to improve them with respect to the quality of the data being collected being sufficient. In this case, this means that the responses from research subjects are the result of intuitions.

4. Bolger, F., & Rowe, G. (2015). The aggregation of expert judgment: do good things come to those who weight?. *Risk Analysis*, 35(1), 5-11.

Keyword(s): article, use

This article argues against what it refers to as the “classical method” by arguing “that no significant benefits are likely to accrue from unequal weighting in mathematical aggregation [of expert judgments]” (5). It is stated that the argument relies on the assumption that it is incredibly difficult to construct “reliable and valid measures of substantive expertise upon which to base weights” (5). In addition, the article discusses practical problems that arise in attempts to evaluate experts. It is claimed that the argument generalizes in that it not only demonstrates that the classical method is problematic, but that all “externally imposed unequal weighting schemes” (5) are problematic. It cites empirical studies and other arguments found in other articles in support of its claims, both in addition to, and in support of, its original arguments.

5. Bolger, F., & Wright, G. (1994). Assessing the quality of expert judgment: Issues and analysis. *Decision Support Systems*, 11(1), 1–24. [http://dx.doi.org/10.1016/0167-9236\(94\)90061-2](http://dx.doi.org/10.1016/0167-9236(94)90061-2)

Keyword(s): article, definition, empirical studies, modeling

This paper “focus[es] on the issue of identifying the factors contributing to expertise and propose[s] some theoretical constructs relating to the conditions under which valid expert judgment .... [is] demonstrated” (2). It begins by examining “the approaches taken towards the definition of expertise by writers in various fields” (2). It is concluded that the two identified approaches, the professionalism approach and the competence approach, are both problematic and are subject to many of the same criticisms.

Following that, it assesses relevant empirical studies of expert judgment and decision making. It is discovered that the research findings are contradictory in that some found expert to be accurate decision makers while others found that they were not. The first of two reasons proposed to account for sub-optimal expert performance are experts' being forced to work outside of their domain(s) of expertise or “express their judgments in unfamiliar metrics” (21). The second is that the unavailability of objective data and models as well as usable feedback may make it difficult for experts to improve their performance significantly over time with experience.

Finally, it “conclude[s] that judgmental performance is largely a function of the interaction between [the]... two dimensions of ecological validity and learnability - if both are high then good performance will be manifest but if one or both are low then performance will be poor” (22). From this, Bolger and Wright also infer that experts “asked to assess subjective probabilities for the occurrence of novel future events” are likely to exhibit relatively poor performance (22). Their conclusion was retroactively tested on a sample of studies and those two dimensions were determined “to be reasonably good predictors of the performance outcomes of these studies” (22).

6. Burgman, M., Fidler, F., McBride, M., Walshe, T., & Wintle, B. (2006). Eliciting expert judgments: literature review.

Keyword(s): definition, elicitation, literature review, modeling, probability

This is a fairly comprehensive literature review on the elicitation of expert judgments. It “reviews methods for eliciting probabilities, quantities, and conceptual models[,] ... defines ‘expert’ status, reviews the literature on biases and heuristics in expert judgements, ... outlines methods for detecting and eliciting values, attitudes and motivations, ... describes direct and indirect techniques for eliciting point estimates and uncertainties for quantities, frequencies and probabilities, and for eliciting the structure of conceptual models[,] ... evaluates the use of language-based risk categories[,] and describes methods to detect and adjust for bias and variability in expert judgements” (1).

Additionally, it makes conclusions about what it takes to be “useful additions to elicitation protocols” (1) as a result of reviewing the literature on expert judgment elicitation available before the time of its publication in 2006. Two such additions are the uses of both feedback and training to elicitation protocols. It is argued that these techniques will improve the accuracy of

expert judgments among other things. The review also suggests that it is worthwhile to implement the formal techniques it discusses for elicitation, calibration, and verification as they appear to be reasonable and their efficacy in practice has not been ascertained. The empirical studies used to assess these techniques would then allegedly allow us to determine which of them are the most promising.

7. Camerer, C. F., & Johnson, E. J. (1997). 10 The process-performance paradox in expert judgment: How can experts know so much and predict so badly? *Research on Judgment and Decision Making: Currents, Connections, and Controversies*, 342.

Keyword(s): book chapter, empirical studies, formation

This chapter sets out to answer the question “How can experts know so much and predict so badly?” which it calls “the process-performance paradox in expert judgment.” It begins by noting two common views on expert judgment. The first is that experts are not better predictors than those with less expertise and the second is that expertise may be developed “only after much instruction, practice, and experience” (342-343). It adduces results from various studies in support for the former. It goes on to suggest that, while experts “are successful at generating hypotheses and inducing complex decision rules... their knowledge and rules have little impact on... [their] performance” (357). It takes one important implication of this to be that the actual performance of experts should not be overlooked by only attending to factors related to one's expertise or expert status.

8. Carol-anne, E. M., Regehr, G., Mylopoulos, M., & MacRae, H. M. (2007). Slowing down when you should: a new model of expert judgment. *Academic Medicine*, 82(10), S109–S116.

Keyword(s): article, definition, formation, modeling

This article argues that expertise involves not merely experience with analytic processes and/or nonanalytic resources such as heuristics, but the ability to “coordinate these resources, using efficient nonanalytic processes for many tasks, but transitioning to more effortful analytic processing when necessary” (S109). It proposes a model of expert judgment described as “slowing down when you should” (S109). Special attention is given to experts' situational awareness and the attention and effort devoted to the various tasks they perform in order to describe how and why experts transition between analytic and nonanalytic ways of thinking. “Through experience, increasingly complicated and complex activities become automated, and, therefore, cognitive resources are freed up for engagement in other activities” (S112-S113). Keeping an intentional vigil on the environment as these nonanalytic process function is what allows the expert to recognize the need to transition back to more effortful analytic processes when the situation warrants it. This article also argues that expertise is best recognized from a process perspective rather than from an achievement perspective – i.e., having achieved a certain status. It is then claimed that this conception helps to make sense of how those not of expert status can, at times, exhibit expertise while those who are considered to be experts can, at times, fail to do so.

9. Cooke, R. M., & Goossens, L. L. H. J. (2008). TU Delft expert judgment data base. *Reliability Engineering & System Safety*, 93(5), 657–674.  
<http://dx.doi.org/10.1016/j.res.2007.03.005>

Keyword(s): article, modeling, use

This article begins by reviewing the classical model of structured expert judgment in order to analyze applications of expert judgment with respect to the resulting accuracy of performance-based weighing vs. equal weighing. In performance-based weighing, an expert's judgment is weighted proportionally to the accuracy of that expert's past judgments. In equal weighing, all expert's judgments are assigned equal weights. Accuracy is assessed via the use of seed variables in conjunction with an expert's judgment. The two are compared post-hoc and their proximity to one another provides a measurement of the accuracy of a particular judgment. It ultimately concludes that, while preferable to equal weighing performance-based weighing is still far from ideal. It goes on to argue that “our models must be quantified with uncertainty distributions, rather than 'nominal values' of undetermined pedigree” (672).

10. Cooke, R. M., & Probst, K. N. (2006). *Highlights of the expert judgment policy symposium and technical workshop*. Resources for the Future Washington, DC.  
<http://www.rff.org/files/sharepoint/Documents/Conference-Summary.pdf>

Keyword(s): conference proceedings, elicitation, empirical studies, modeling, probability, selection, use

These conference proceedings deal with expert judgment and related topics. Granger Morgan gives a presentation entitled “Regulation under Uncertainty” which “focused on the characterization of uncertainty and decisionmaking under uncertainty, and gave some summary guidance on reporting and analyzing uncertainty” (3). “The bottom line was: 'Without at least some quantification, qualitative descriptions of uncertainty convey little, if any, useful information” (3).

The proceedings go on to discuss some recent applications of expert judgment, one of which is the use of a “formalized procedure for providing scientific advice using the software system EXCALIBUR @TUDelft” with respect to matters involving volcanoes such as the assignment of an alert level to a situation and the giving of evacuation advice.

Another is expert judgment regarding the health effects of fine particulate matter PM<sub>2.5</sub>. The latter is explored in more detail in a talk given by John Evans entitled “What Experts Know about PM Risks” which focuses on the 1991 Kuwaiti oil fires and their dose-response relationship to health effects. The effected population's age-structure, the composition of the smoke caused by the fire, and the exposure level and pattern are all taken into account. An ACS study concluded that the fires resulted in around 35 deaths, while expert judgments weighted equally with one another predicted 82 deaths and those same judgments assigned performance-based weights resulted in an average judgment of 35 deaths.

This talk is followed by one given by Katherine Walker entitled “Expert Judgment about Uncertainty in PM<sub>2.5</sub> Mortality: What Have We Learned?” in which she describes her work with others on a “project to develop a more comprehensive characterization of uncertainty in the

concentration-response relationship between PM<sub>2.5</sub> and mortality for use in regulatory impact assessments for proposed regulations” (13). A pilot study was completed, but soon after, a more comprehensive study began (the results of which are not linked to in the conference proceedings).

Next is an account of the round table on practical issues of expert judgment that was held at the conference. “[T]here was substantial consensus regarding the desirability of using structured expert judgment to quantify uncertainty, particularly where the stakes are high” (14). It was generally agreed upon that having 6-12 experts for an instance of expert judgment is typically sufficient. Some touted the benefits of including experts with minority viewpoints among the panel of experts involved in a particular judgment scenario. Preferences were expressed for privacy with respect to the association of experts' names with their judgments.

The conference proceedings go on to present a list of challenges for policymakers and practitioners. The process of expert elicitation is discussed in terms of how it might best be conducted. Also, the need for the quantification of uncertainty and a proper method of obtaining it is reiterated. This presentation “ended with a call for more research devoted to communicating with decisionmakers and decisionmaker training in understanding uncertainty” (19).

Excluding the account of how the technical workshop held at the conference went, the concluding section of the conference proceedings is a list of key issues addressed at the conference. These issues include: motivational bias; the costs of expert judgment studies; individual vs. group elicitation; the choice of whether to aggregate judgments; the choice of method in how to aggregate judgments if it is decided that they are to be aggregated; whether it is better to elicit the input or the output of a model; whether an experiment's performance being validated on a series of seed variables means that that expert will do well on the variables of interest; whether there have been cases of post hoc direct verification of experts' performance; whether there are any demographic aspects that can determine experts' performance; the connection between expert elicitation and peer review; how to determine whether a problem is an expert judgment problem; and where more research effort in expert elicitation could be spent (20-23).

11. Egilmez, G., Gumus, S., & Kucukvar, M. (2015). Environmental sustainability benchmarking of the U.S. and Canada metropolises: An expert judgment-based multi-criteria decision making approach. *Cities*, 42, Part A, 31–41.  
<http://dx.doi.org/10.1016/j.cities.2014.08.006>

Keyword(s): article, formal methods, probability

This article presents a “four-step hierarchical fuzzy multi-criteria decision-making approach” to the issue of “environmental sustainability performance assessment” (31). The first step of the approach is to determine the various sustainability performance indicators, collect data, and contact the relevant experts. “In the second step, experts are contacted...; sustainability performance evaluation forms are delivered; and then expert judgment results are obtained and quantified, respectively” (31). In the third step, the “Multi-criteria Intuitionistic Fuzzy Decision Making model is developed” and subsequently applied to the relevant data. Finally, “the sustainability scores and rankings of the 27 metropolises [located in either the U.S. or Canada], results analysis and discussions, and statistical highlights about the research findings are

provided” (31).

Most relevant to the issue of expert judgment is this article's discussion of Intuitionistic Fuzzy Decision Making. This kind of decision-making has “been applied to environmental sustainability problems where multiple alternatives need to be benchmarked or ranked” (33). The article mentions a few approaches to this kind of decision making that have been used by others before introducing the approach it uses. It notes that the difference between its approach and past approaches is that it aims “to combine IFS [(Intuitionistic Fuzzy Set)] methodology with MCDM [(Multi-criteria Decision Making)] framework in... the two stages [of] expert judgment and ranking” (33).

The article goes on to suggest that its preferred approach has several merits. The first is that “such a specific language approach makes it possible for experts to complete the survey easily and reflects the fuzziness or uncertainty associated with the importance of selected sustainability indicators” (33). The next two have to do with the transformation of selected sustainability indicators into a fuzzy dataset. It argues that this leads to a compatibility in the types of data dealt with which in turn results in a more robust assessment of uncertainty. It also claims that it allows for fair comparison between the cities because “[u]sing linguistic terms prevents the deteriorating effects of scale differences in numerical data on the final ranks of the cities” (34).

12. Einhorn, H. J. (1974). Expert judgment: Some necessary conditions and an example. *Journal of Applied Psychology*, 59(5), 562-571. <http://psycnet.apa.org/journals/apl/59/5/562/>

Keyword(s): article, empirical studies, formation

This article proposes the following three criteria “as being indicative of expert judgment: (a) Experts should tend to cluster variables in the same way when identifying and organizing cues. (b) Expert judgment should be highly reliable (intrajudge reliability), show both convergent and discriminant validity, and be relatively free of judgmental bias when measuring cues, (c) Experts should weight and combine information in similar ways” (562). A study was done involving three medical pathologists presumed to be of expert status. In addition to a global judgment, the subjects were asked to weigh various pre-selected factors and report their estimates. The results of the study, for the most part, indicate that the three proposed criteria are fairly appropriate. The most problematic one in light of the results was the requirement that experts weight and combine variables similarly. It is proposed that this might be due to the presence of other, unnecessary conditions that are indicative of expert judgment. A few criticisms such as the requirement that there be interjudge agreement are addressed. While not entirely dismissed, they are considered to not pose any significant problems for the propriety of the proposed criteria.

13. Ericsson, K. A. (2014). Why expert performance is special and cannot be extrapolated from studies of performance in the general population: A response to criticisms. *Intelligence*, 45, 81-103.

Keyword(s): article, empirical studies, formation

While not directly dealing with the expert judgment process, this article does discuss the



issue of expert performance. A better understanding of expert performance might yield insights into how the elicitation process in the process of expert judgment might best be constructed, among other things.

The article argues that “the assumption that data on large samples of beginners can be extrapolated to samples of elite and expert performers” (81). It identifies the latter sort of approach as the individual-difference approach and contrasts it with the expert-performance approach it goes on to present. It is claimed that this sort of specialized approach is needed for the study of expert-performance. In order to formulate “criteria for reproducible objective expert performance and acceptable methodologies for collecting valid data” (81) it is suggested that “studies of the acquisition of expert memory with detailed experimental analysis of the media mechanisms” be undertaken.

14. Ericsson, K. A., & Smith, J. (1991). *Toward a general theory of expertise: Prospects and limits*. Cambridge University Press.

Keyword(s): book, empirical studies, formation, modeling

This book discusses the idea of a general theory of expertise. In doing so, numerous promising opportunities for future research and obstacles to be expected in conducting such research are addressed. In trying to discover general characteristics of expert performance across different domains, it is argued that there's a difficulty integrating the results of studies from different domains. Not only do experts typically specialize in certain domains, but it is claimed that researchers of their performance are often similar in this respect. It is suggested that this warrants an approach to identifying general characteristics of expert performance that focuses on integrating findings from various domains in order to identify general characteristics of expertise.

15. Ettenson, R., Shanteau, J., & JACKKROGSTAD. (1987). Expert judgment: Is more information better? *Psychological Reports*, 60(1), 227–238.  
<http://www.amsciepub.com/doi/pdf/10.2466/pr0.1987.60.1.227>

Keyword(s): article, empirical studies, formation

This article first discusses with how experts deal with (different amounts of) information and goes on to claim, against what is suggested is the popular view--that “experts are limited in their ability to process information--that experts non-use of certain information is better explained as their dismissal of it as less relevant than other, more relevant information.

A study designed to test this article's main claim was devised and carried out, and a discussion of that study is present in this article. “The participants' task in this study was to determine the degree of materiality in 32 hypothetical auditing cases. Three groups of auditors were asked to evaluate cases described by eight auditing cues” (228). One aspect of the results--the student group's use of a similar quantity of information as the expert group--was consistent with this article's contention and the claim it was opposed to. However, it is argued that other aspects of the results--the significantly greater consistency among experts' judgments than those of students and experts' primary reliance on a smaller subset factor than students in their

decision-making processes--is taken as evidence for the truth of this article's primary claim. The idea is that experts, through the accumulation of experience in their domain of expertise, develop similar decision-making strategies and that one important part of those strategies is the classification of different kinds of information as more or less relevant.

16. Evans, J. S., Gray, G. M., Sielken, R. L., Smith, A. E., Valdezflores, C., & Graham, J. D. (1994). Use of Probabilistic Expert Judgment in Uncertainty Analysis of Carcinogenic Potency. *Regulatory Toxicology and Pharmacology*, 20(1), 15–36.  
<http://dx.doi.org/10.1006/rtp.1994.1034>

Keyword(s): article, elicitation, empirical studies, formation, selection, use

This article's primary objective is to evaluate two methods of determining carcinogenic potency--risk distributions of experts and EPA risk calculations. However, the study is a relatively principled and rigorous instance of the process of expert judgment which exemplifies many of the elements deemed essential to a properly conducted instance of such a process by recent expert judgment research and literature. The four phases of selection, elicitation, formation, and use are all present in this instance, and many of the important considerations mentioned in research on expert judgment are taken account of and applied in it as well.

17. Gray, A. R., MacDonell, S. G., & Shepperd, M. J. (1999). Factors systematically associated with errors in subjective estimates of software development effort: the stability of expert judgment. In *Software Metrics Symposium, 1999. Proceedings. Sixth International* (pp. 216–227). [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=809743&tag=1](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=809743&tag=1)

Keyword(s): article, empirical studies, heuristics and biases

This article seeks to identify factors that commonly lead to experts' errors in their judgments in the process of software development and to construct models that capture the effects those factors have on experts' judgments. It discusses various sources of those inaccurate estimations such as: misunderstood or not-fully-understood recent changes in technology, lack of experience in the relevant domain, lack of experience with the relevant system (or a component of it), other relevant influences in the expert's background, and reasons having to do with political or motivational goals.

It is suggested that, since these errors are systematic, they might be corrected for or eliminated if addressed appropriately. Various models are analyzed with respect to their efficacy in determining “the direction and magnitude of any systematic biasing effect” (218) that experts' judgments are susceptible to. Two available approaches for the task of how the negative influence of such effects might be addressed are discussed. One might approach this task by “informing project managers [(experts)] that they are making such errors or by automatically correcting estimates in such circumstances” (218).

18. Gordon, T. (n.d.). Carneades 3.7 User Manual. Retrieved from <https://carneades.github.io/manuals/Carneades3.7/carneades-3.7-manual.pdf>

Keyword(s): argumentation

This is the user manual for version 3.7 of the Carneades argumentation system software. The program is open-source and is used to model arguments. A significant benefit the tool offers is an integrated inference engine for constructing arguments from knowledge-bases. In the safety engineering community, rigorous argument has been used as a method of attaining assurance that processes of expert judgment have been properly constructed and conducted, so a software tool capable of modeling such arguments is certainly relevant to their purposes. The latest version of the program is 4.2, but, as of the time this bibliography was published, no comprehensive user manual has been created for that version.

19. Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Transactions on Systems, Man, and Cybernetics*, 17(5), 753–770. <http://doi.org/10.1109/TSMC.1987.6499282>

Keyword(s): article, empirical studies, formation

This article discusses the results of direct comparison--as opposed to the usual method of assessment used in similar studies, indirect comparison--between the efficacy of intuitive and analytical cognition in expert judgment. It begins with a compelling argument for the advantages direct comparison has over indirect comparison with respect to the practice of expert judgment. The idea is that the former is more representative of the actual practice of expert judgment and that this makes such a comparison more useful in terms of allowing one to draw conclusions about that practice.

The results of the study performed yield the surprising result of intuitive and quasi-rational cognition frequently outperforming analytical cognition in terms of the empirical accuracy of judgments. These results conflict with similar studies that use indirect comparison; while those studies would lead one to think that analytical approaches such as following an algorithm are preferable for an expert to use whenever they are available, this one suggests that the proper decision of what kind of cognition experts should use in such situations should involve careful consideration of the judgment instance.

After enumerating additional, significant conclusions drawn from the study, the article concludes with the suggestions that “further research on the differential effects of task properties will markedly increase our ability to predict the form of a person's cognitive activity from a knowledge of task properties” and that “the relation between the two should predict [expert] performance” (766).

20. Jorgensen, M. (2005). Practical guidelines for expert-judgment-based software effort estimation. *IEEE Software*, 22(3), 57–63. <http://doi.org/10.1109/MS.2005.73>

Keyword(s): article, empirical studies, formation, probability, selection

This article gives “seven guidelines for producing realistic software development effort estimates. The guidelines derive from industrial experience and empirical studies” (57). They differ from other guidelines in that: “they base estimates on expert judgments rather than models, [...] they are easy to implement,” and “they use the most recent findings regarding judgment-based effort estimation” (57). The seven proposed guidelines are: “Don’t mix estimation, planning, and bidding” (58), “Combine estimation methods” (58-59), “Ask for justification” (60), “Select estimation experts with experience from similar projects” (60-61). “Accept and assess the uncertainty of effort usage” (61), “Provide learning opportunities” (61-62), and “Consider postponing or avoiding effort estimation” (62-63). Helpful methods of implementing the above guidelines are presented. Studies and observed tendencies are cited often to provide support for claims.

21. Joyce, E. J. (1976). Expert Judgment in Audit Program Planning. *Journal of Accounting Research*, 14, 29–60. <http://www.jstor.org/stable/2490445>

Keyword(s): article, empirical studies

This article seeks to determine whether or not auditors make valid judgments. It is stated that empirical studies have been inconclusive in that they have shown “substantial differences among the judgments of individual auditors” (29). Difficulties associated with determining what constitutes correct and incorrect judgments are discussed and it is argued that consensus among experts is a plausible indicator of the former (30-31). A research strategy is outlined and argued for which involves an experiment that asks subjects to perform one of several systematically different audit planning tasks (34-35). It is suggested that the way in which expertise is typically acquired in a dynamic, probabilistic environment in which the auditor rarely gets exact feedback and then claimed that this should lead us to expect individual differences amongst auditors (36-37).

The article goes on to discuss the study conducted for the purpose of assessing the validity of auditors' judgments. The study is described as “an extension of Ashton's (1974) work on internal control evaluation” (37). The results were fairly consistent with previous research in that auditors exhibited a significant degree of individual difference among their judgments. Implications of the study are then discussed. The article claims that its results do not “imply that additional steps should be taken to eliminate the differences [between auditors' judgments]. The appropriate course of action depends on the costs of such differences and the costs of reducing them” (54). It goes on to discuss the following alleged limitations of the study performed: statistic inferences cannot be made from its results “since the subjects were not randomly selected from the population of auditors or from firms” (54), its non-representative nature that did not take into account the nonzero intercorrelations that probably exist among the independent variables in nature” (55), its tasks were sent to participants via mail and it was not administered and supervised at firm offices, and it dealt solely with individual audit judgment while the subjects in the study were usually subject to review by superiors (55).

The following further areas of research are suggested at the end of the article: “research into individual differences among auditors' judgments... using representative designs where possible” (55-56), “research into the dynamics of the audit review process” (56), an attempt to model auditors using something referred to as AN-OVA, a study to “examine the question of

whether individual differences in judgment at the audit-program-planning level lead to differences in the audit opinion rendered and/or the cost of the audit” (56), and research on the use of decision heuristics in auditing tasks (56-57).

22. Kandlikar, M., Ramachandran, G., Maynard, A., Murdock, B., & Toscano, W. A. (2006). Health risk assessment for nanoparticles: A case for using expert judgment. In *Nanotechnology and occupational health* (pp. 137-156). Springer Netherlands.

Keyword(s): article, elicitation, use

This article begins by explaining why uncertainty in risk assessments in nanoparticle exposure is relatively extreme. An approach is outlined in which the “degree of consensus and/or disagreement between experts on different parts of the exposure-response paradigm” are used (137). Judgments are elicited “from a wide range of experts on different parts of the risk causal chain” (137). Additional applications of the proposed approach such as research prioritization and budget allocation are discussed. It is suggested that this approach ideally be repeated over the course of several years as “the state of scientific knowledge will... improve and uncertainties may possibly reduce” (137).

23. Keeney, R. L., & von Winterfeldt, D. (1989). On the uses of expert judgment on complex technical problems. *IEEE Transactions on Engineering Management*, 36(2), 83–86.  
<http://doi.org/10.1109/17.18821>

Keywords(s): article, elicitation, formation, use, probability

This article discusses a range of issues including the role and uses of expert judgment in “analysis of technical problems, improvement of expert judgments, and interpretation of judgements in analysis” (83). It argues that there is significant value in “quantifying expert judgements to complement the expert's qualitative thinking and reasoning” (83). It also discusses “the relationships between procedures to quantify judgments and the general principles of engineering” (83). It asserts that the competency of all involved with the probability elicitation process is an important factor in accurately quantifying expert judgments (84). Similar stress is placed on documenting the process of expert elicitation and ensuring that expert judgment is not misused (85-86). It concludes by discussing various considerations relevant to instances of expert judgment that involve multiple experts (86).

24. Meyer, M. A., & Booker, J. M. (2001). *Eliciting and analyzing expert judgment: a practical guide* (Vol. 7). SIAM.

Keyword(s): book, elicitation, use

This book discusses various techniques for eliciting and analyzing expert judgment. It makes a case for the essential role expert judgment plays in various kinds of assessment for which measurements are test results are not available. It also addresses how one might develop

elicitation methods and tools suitable for the needs of particular situations in a particular domain.

25. Lin, S. W., & Bier, V. M. (2008). A study of expert overconfidence. *Reliability Engineering & System Safety*, 93(5), 711-721.

Keyword(s): article, empirical studies, formation

This article studies what it considers to be “one of the most common (and potentially severe) problems in expert judgment” (711)—expert overconfidence. An expert who has an inappropriate level of confidence in their judgments is arguably more prone to making inaccurate judgments than experts with the level of confidence that is appropriate to the judgment context they are dealing with, so insight into what causes expert overconfidence and how it might be mitigated is of interest to the community which uses the expert judgment process to make important decisions.

In order to “assess the extent of expert overconfidence, [the article] analyze[s] a large data set on expert opinion compiled by Cooke and colleagues at the Technical University of Delft and elsewhere” (711). It is noted that “[s]ignificant differences in the extent of overconfidence were found among studies, among experts, and among questions within a study” (711). It is claimed that the results of the “analysis suggest that much of the apparently question effect may be due to noise rather than systematic differences in the difficulty of achieving good calibration for different questions” (711). Finally, it is argued that “the results support the differential weighting of experts, since there are significant differences in expert calibration within studies” (711).

26. Mizrahi, M. (2013). Why arguments from expert opinion are weak arguments. *Informal Logic*, 33(1), 57–79.

[http://ojs.uwindsor.ca/ojs/leddy/index.php/informal\\_logic/article/view/3656](http://ojs.uwindsor.ca/ojs/leddy/index.php/informal_logic/article/view/3656)

Keyword(s): argumentation, article, definition, empirical studies, formation

This articles claims that arguments from expert opinion are weak arguments, and that appeals to expert opinion are appeals to authority. Expert opinion/judgment, in the context of assurance cases, refers to the cognitive authority of the expert. Experts may possess administrative authority in addition to cognitive authority. For example, a head surgeon may, in addition to being incredibly knowledgeable in his or her particular field, also be responsible for making decisions and managing other, “lower-ranked” employees. This second aspect, or the “dominance” and ability to impose their will on others is not considered to be relevant in this article and other discussions of expert opinion unless it is mentioned negatively with respect to its tendency to act as a biasing factor

Additionally, in arguments, expert opinion is usually used to talk about the cognitive authority of a particular expert rather than the collective consensus on a topic by a group of experts on the subject that topic belongs to.

By “weak argument,” the author means that expert opinions/judgment are only slightly more likely to be accurate (by a statistically insignificant margin of 5% or less) than the

opinions/judgments of a non-expert about the same topic. This claim is based on experiments in the emerging field of Experimental Philosophy, where empirical data is examined and analyzed to reach conclusions, much like it is in the various natural sciences. In these experiments, the opinions/judgments were made in situations in which the experts were essentially on “equal footing” with the non-experts. In other words, there were no complicated procedures that only the experts would know how to perform that would disadvantage the non-experts. This being the case, these experiments were allegedly tests of to what degree, if any, expert intuitions are more accurate than those of non-experts.

27. Ortiz, N. R., Wheeler, T. A., Breeding, R. J., Hora, S., Meyer, M. A., & Keeney, R. L. (1991). Use of expert judgment in NUREG-1150. *Nuclear Engineering and Design*, 126(3), 313–331. [http://dx.doi.org/10.1016/0029-5493\(91\)90023-B](http://dx.doi.org/10.1016/0029-5493(91)90023-B)

Keyword(s): article, definition, elicitation, selection, use

This article discusses the expert judgment process used in the assessment of five US nuclear plants. The phases of selection, elicitation, and use are dealt with in detail. Of particular interest are the careful selection of issues for expert judgment and the elicitation training used to prepare for the actual process. It summarizes one issue of the relevant expert judgment process using accident frequency analysis and another using accident progression analysis. It concludes with suggestions for ways in which the use of explicit expert judgment might be improved with respect to its use in addressing complex technical issues (313).

It recommends an expert judgment process consisting in a sequence of the following six steps: “(1) selection of issues and experts; (2) elicitation training; (3) presentation of issues to the experts; (4) preparation of issue analyses by the experts; (5) discussion of issue analyses and elicitation of expert's judgments, and (6) recomposition and aggregation of expert judgments” (315). With respect to the first step, it suggests that two selection processes take place concurrently with them both in mind. The elicitation training “familiarized the experts with the overall project and demonstrated how their judgments would be used in the analysis, [...] provided debiasing training, [...] and] afforded practice making probability assessments and decomposing complex issues into smaller subissues” (316).

28. Otway, H., & von Winterfeldt, D. (1992). Expert Judgment in Risk Analysis and Management: Process, Context, and Pitfalls. *Risk Analysis*, 12(1), 83–93. <http://doi.org/10.1111/j.1539-6924.1992.tb01310.x>

Keyword(s): article, empirical studies, selection, elicitation, use

This article discusses expert judgment's uses in risk assessment and management, how it might best be structured for different kinds of instances, and solutions for preventing commonly observed problems associated with its use from occurring. It is claimed that expert judgment fills an essential role unable to be properly filled by a process consisting of strict adherence to scientific practice. Its purpose is stated as furthering the ongoing effort taken to formalize expert judgment. The article identifies four categories of expert judgment and presents three case

studies meant to exemplify some of its common pitfalls. In light of its findings, it claims that expert judgment will become more heavily relied upon in risk assessment and management as time goes on and its documentation enhanced in terms of its openness to the public and its detail. Finally, it is suggested that this has implications for increased public involvement in scientific and technical affairs.

29. Peng, W., Zan, M., & Yi, T. (2011). Application of Expert Judgment Method in the Aircraft Wiring Risk Assessment. *Procedia Engineering*, 17, 440-445.

Keyword(s): article, empirical studies, formation, probability, use

This article deals with the FAA's "risk assessment means for [its] electrical wiring interconnection system, in which expert judgment plays an important role" and "discusses and reviews the evaluation method for the expert judgment from the two aspects: the validation of each expert judgment and the coefficient of agreement as a group" (440). This article is interesting in that it takes a relatively critical approach towards the use of expert judgment. In the context the article discusses, "the huge number of complex judgment[s one or more experts must make] will lead to high probability that the experts make the wrong choice" and it is argued that "[a]n evaluation and validation of the expert judgment before using the data is necessary" (440). It is claimed that the method the article proposes allows one to "find out if each expert is specifying a true preference structure in his/her answers or just assigning answers in a random fashion, but it cannot exactly check if the judgment of expert is accurate" (444).

30. Roman, H. A., Walker, K. D., Walsh, T. L., Conner, L., Richmond, H. M., Hubbell, B. J., & Kinney, P. L. (2008). Expert Judgment Assessment of the Mortality Impact of Changes in Ambient Fine Particulate Matter in the U.S. *Environmental Science & Technology*, 42(7), 2268–2274. <http://doi.org/10.1021/es0713882>

Keyword(s): article, empirical studies, elicitation, formation, probability, selection

This article presents findings of an "expert judgment study that comprehensively characterizes uncertainty in estimates of mortality reductions associated with decreases in fine particulate matter (PM<sub>2.5</sub>) in the U.S" (2268). The Environment Protection Agency (EPA) "has applied state-of-the-art expert judgment elicitation techniques to develop probabilistic uncertainty distributions that reflect the broader array of uncertainties in the C-R [(concentration-response)] relationship" (2268).

Of particular interest is this article's discussion of the highly-structured process of expert elicitation used in the study. The experts attended a pre-elicitation workshop that was "designed to introduce the project, familiarize them with expert judgment and the elicitation process, and foster critical discussions of key evidence relevant to the questions posed by the study" (2270).

The next step of the elicitation process was the elicitation interview. "The elicitation team included two interviewers, one experienced in the elicitation of expert judgments... and one with expertise in PM health effects and exposure" (2270). During the interviews, "experts were asked to think systematically about, and cite, the evidence in support of their responses" (2270). "Each



expert was also given the opportunity to participate in Internet-based conferencing with Industrial Economics representatives... who provided real-time graphical and quantitative feedback regarding the expert's C-R function distribution and responses to protocol questions” (2270).

The penultimate step of the elicitation process was a post-elicitation workshop. The purpose of the workshop was “to share results anonymously with the group, highlight areas where expert opinion varied, clarify points of confusion, allow experts to raise issues for discussion, and encourage each expert to critically review his judgments” (2270). The final step took the form of a meeting summary which was sent to the experts who, at that point, “were provided an opportunity to revise their qualitative and/or quantitative judgments privately using a standardized form” (2270).

31. Rosqvist, T., Koskela, M., & Harju, H. (2003). Software quality evaluation based on expert judgement. *Software Quality Journal*, 11(1), 39-55.

Keyword(s): article, elicitation, formal methods, modeling, probability

This article presents a method for using expert judgment in the evaluation of software quality. The underlying principle of the approach is stated to be “the encoding of experts’ tacit knowledge into probabilistic measures associate with the achievement level of software quality attributes” (39). In the next step of the method, “[a]n aggregated quality measure is obtained based on preference statements [by the experts participating in the expert judgment process] related to the quality attributes” (39). The paper describes each component of the method in great detail, and provides support for the inclusion of each. It is argued that merits of the method are the support it offers in “systematic identification of discrepancies in the experts’ judgments... [and] the identification of the most influential expert judgments determining the software evaluation outcome” (51).

32. Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, 53(2), 252–266. [http://dx.doi.org/10.1016/0749-5978\(92\)90064-E](http://dx.doi.org/10.1016/0749-5978(92)90064-E)

Keyword(s): article, definition, empirical studies

This article seeks to reconcile two contrasting views on experts seen in the literature. Judgment and decision research has portrayed them as susceptible to biases. On the other hand, cognitive research has portrayed them as rather more competent and more effective than novices in nearly all aspects of cognition. The article calls its alternative, newly proposed approach “the Theory of Expert Competence’, which assumes [that] competence depends on five components: (1) a sufficient knowledge of the domain, (2) the psychological traits associated with experts, (3) the cognitive skills necessary to make tough decisions, (4) the ability to use appropriate decision strategies, and (5) a task with suitable characteristics” (252).

Developing this theory is the focus of the paper. The authors view the attention given to task and domain characteristics in prior research as insufficient. It suggests that differences in results

between the two aforementioned research fields can be explained by the differences in those characteristics.

It is asserted that the latter finding has several important implications. It notes that it may contribute to an improved understanding of when expert systems are preferable to actual experts or vice-versa. The following four conclusions are reached: “[ (1) ] expertise must be looked at from the perspective of experts, not as something to be defined within the constraints of available hardware and software; [ (2) ] experts cannot be expected to explain everything about what they do; [ (3) ] more emphasis should be placed on the traits, skills, and strategies of human experts when building computer systems; [ (4) ] different types of expert systems may be needed to reflect left-side[-brain] and right-side[-brain] expertise.” (262-263).

33. Shanteau, J., & Stewart, T. R. (1992). Why study expert decision making? Some historical perspectives and comments. *Organizational Behavior and Human Decision Processes*, 53(2), 95–106. [http://dx.doi.org/10.1016/0749-5978\(92\)90057-E](http://dx.doi.org/10.1016/0749-5978(92)90057-E)

Keyword(s): article, definition, empirical studies, heuristics and biases, modeling, formation, probability

This article makes a case for the importance of studying expert decision making. It defines expertise in terms of the skill an expert has at making accurate judgments gained through experience and/or training. It consists of three main sections. “The first summarizes previous research in the J/DM [(Judgment/Decision Making)] field. The second contains observations about why this line of research is important. The final section comments on the nine papers in this special issue” (95-96)

The first section begins with a discussion of psychometric research on expert judgment designed to test the validity and reliability of their judgments. The general conclusion of such research “is that experts are lacking in validity and reliability and that more information increases confidence but not accuracy” (97). The general conclusions of linear model analysis have been “that relatively few cues account for virtually all of the systematic variance [in human judgment]” and that “[s]ince the expectation of greater information use by experts was not supported, [...] experts are limited in the same manner as novices” (97-98). The research on biases and heuristics as well as probability assessments have had similar results in that they too “paint a dismal picture of the abilities of experts” (99). The section finishes with a discussion of a studies that have depicted experts favorably in terms of the accuracy of their judgments. It is suggested these studies have led to “the emerging view that some experts can make competent decisions in at least some situations” (100).

The second section presents three reasons why research on experts is valuable. “The first involves the generalizability of results from naive subjects to experts. The second reflects the increasing importance of expert systems. The third reason is the recognition that experts are interesting in their own right” (100).

The third section is an introduction to the selection of papers chosen in the issue of the journal in which this article was published. It states that they have a few common features. “First, they are concerned with when, and where, experts are superior decision makers. [...] Second, the

studies described here illustrate the application of 'second-generation' research methods to expert decision makers. [...] Third, the authors are experienced in conducting research on experts” (103).

34. Stewart, T. R., Roebber, P. J., & Bosart, L. F. (1997). The Importance of the Task in Analyzing Expert Judgment. *Organizational Behavior and Human Decision Processes*, 69(3), 205–219. <http://dx.doi.org/10.1006/obhd.1997.2682>

Keyword(s): article, empirical studies, formation, modeling, use

This article discusses studies that tested “the accuracy of judgmental forecasts of temperature and precipitation...” (205). “In contrast to the findings of many studies of expert judgment and forecasting, forecasts were highly accurate and forecaster agreement was high. Human forecasters performed better than an operational forecasting model and about the same as a linear regression model...” In addition to accuracy, the article claims to yield conclusions about the reliability of expert judgments, expert consensus, the performance of groups of experts vs. that of individual experts, and various issues related to models and modeling. It is noted that “[t]ask predictability was an excellent indicator of forecast accuracy”. Finally, “[i]t is concluded that an understanding of the properties of the task is essential for understanding the accuracy of expert judgment.”

35. Uhlenhuth, E. H., Balter, M. B., Ban, T. A., & Yang, K. (1999). International study of expert judgment on therapeutic use of benzodiazepines and other psychotherapeutic medications: VI. Trends in recommendations for the pharmacotherapy of anxiety disorders, 1992–1997. *Depression and Anxiety*, 9(3), 107–116. [http://doi.org/10.1002/\(SICI\)1520-6394\(1999\)9:3<107::AID-DA2>3.0.CO;2-T](http://doi.org/10.1002/(SICI)1520-6394(1999)9:3<107::AID-DA2>3.0.CO;2-T)

Keyword(s): article, empirical studies, formation, use

This article, while it does deal with the topic of expert judgment, does not critically discuss the practice or process of expert judgment itself. Instead, it collects and integrates expert clinical experience and judgment in order to investigate a medical issue. The article demonstrates the utility in using expert knowledge, accurate expert judgments, and the thought processes that lead to those judgments to formulate general strategies for approaching judgment in the medical field.

36. van Noordwijk, J. M., Dekker, A., Cooke, R. M., & Mazzuchi, T. A. (1992). Expert judgment in maintenance optimization. *IEEE Transactions on Reliability*, 41(3), 427–432. <http://doi.org/10.1109/24.159813>

Keyword(s): article, elicitation, modeling, selection, use

This article discusses the use of expert judgment in the domain of maintenance optimization and proposes “a comprehensive method for the use of expert opinion for obtaining lifetime distributions required for maintenance optimization” (427). “The method was motivated by the

practical circumstances governing its implementation. In particular, by the lack of statistical training of the experts and the high demands on their time. The use of a discretized life distribution provides more flexibility, is more comprehensible by the experts in the elicitation stage, and greatly reduces the computation in the combination and updating stages” (427). In addition to discussion of the above method, considerations relevant to the process of the selection of participating experts is discussed in terms of an analysis of expertise.

37. Wagemans, J. H. (2011). The assessment of argumentation from expert opinion. *Argumentation*, 25(3), 329-339.

Keyword(s): argumentation, article, formal methods, modeling

This article “develop[s] a comprehensive tool for the reconstruction and evaluation of argumentation from expert opinion” (329). It combines two accounts—Walton’s account of ‘appeal to expert opinion’ and the pragma-dialectical account of ‘argument from authority’—and argues that doing so allows it to incorporate the merits of both and mitigate the flaws each has individually. More specifically, “[t]he tool is developed by incorporating Walton’s critical questions into a pragma-dialectical framework” (329).

After summarizing each of the above-mentioned accounts, it reviews critiques and criticisms alleged against each one before going on to present its comprehensive tool for assessing arguments from expert opinion. The general pragma-dialectical framework is described and then details about how Walton’s critical questions may be implemented within the framework are presented. It concludes with some considerations that should be taken in to account in the use of the tool presented insofar as how the propriety of it being used in a particular context and how its proper use in such a context may be assessed.

38. Walton, D. (2014). A dialectical analysis of the ad baculum fallacy. *Informal Logic*, 34(3), 276-310.

Keyword(s): argumentation, article, formal methods

This article “applies dialectical argumentation structures to the problem of analyzing the ad baculum fallacy” (276). Some might consider the expert judgment process, as it yields a judgment by an expert who is assumed to be more likely to yield accurate judgments than most or all other individuals, to be guilty of this fallacy, so it is of interest to those who use and rely on the expert judgment process to have it vindicated in this respect.

The article claims that “a suspected instance of the [ad baculum] fallacy [must] proceed through three levels of analysis: (1) an inferential level... (2) a speech act level... and (3) a dialectical level” (276). Additionally, it “adds a new type of dialogue called advising dialogue that needs to be applied at the third level” (276). It concludes “that whether a given ad baculum argument should be properly judged to be fallacious or not is dialectical, meaning that it depends on the type of discourse the argument is supposed to be part of” (305). It is suggested that ad baculum arguments have been considered to be fallacious for so long in logic textbooks is that

such arguments are perceived as threats when not all of them are. Finally, it is claimed that the provided analysis of ad baculum arguments “help[s] us not only to evaluate *ad baculum* arguments but also to explain precisely what goes wrong when such an argument is fallacious by pinpointing a group of dialectical failures that can occur” (307).

39. Walton, D., & Koszowy, M. (2014). Two kinds of arguments from authority in the ad verecundiam fallacy. In *8th Conference of the International Society for the Study of Argumentation, Amsterdam, Netherlands*.

Keyword(s): argumentation, article, formal methods, selection

This paper presents “an argumentation scheme for argument from an administrative authority [that] is formulated along with a matching set of critical questions used to evaluate it” (1). This article appears to be particularly relevant to the selection process of expert judgment as it deals heavily with how an expert is to be assessed. It makes a distinction between two importantly different kinds of argument from authority—those based on administrative authority and those based on cognitive authority. It is suggested that critical questions designed to determine whether or not one is an expert in the relevant domain may be used to ascertain whether an alleged expert truly possesses cognitive authority. Near the end of the paper, it is noted that it is quite common for experts to have both of the above-mentioned kinds of authority and that the argumentation scheme presented is capable of accounting for such cases.

40. Walton, D. and Gordon, T. F. (2013) How to formalize informal logic. *OSSA Conference Archive*. Paper 169.  
<http://scholar.uwindsor.ca/ossaarchive/OSSA10/papersandcommentaries/169>

Keyword(s): argumentation, article, formal methods

This paper formulates ten characteristics of informal logic and offers support for the claim that the Carneades Argumentation System can model all of these characteristics within its formal structure. It believes the “weakest link” in its “chain of argumentation” is their “hypothesis that Carneades can be used to model relevance.” It also mentions that “Carneades argument graphs are evaluated in stages of dialogue” and that “there are some in the informal logic community, and very many in the formal logic and epistemology communities, who might disagree that evaluating an argument properly always requires reference to a conversational (dialogue) setting”. Due to the latter, they have chosen to not include dialogues in the paper “as an essential characteristic of informal logic.” The article itself is concise and to the point, with several illustrative graphics.

41. Walton, D., & Gordon, T. F. (2011, May). Modeling critical questions as additional premises. In *Proceedings of the 8th International OSSA Conference, Windsor, Ontario, ed. F. Zenker*.

Keyword(s): argumentation, article, formal methods, modeling

This article shows how critical questions regarding expert judgment processes can be modeled in the Carneades argumentation system as three kinds of premises—ordinary premises, assumptions, and exceptions. Critical questions are questions meant to assess the competency of an expert and the accuracy of their judgment. Such questions might include questions about what field an expert is an expert in, whether an expert is trustworthy, whether an expert’s opinion is based on evidence, etc. The Carneades argumentation system is an argumentation framework with associated software designed to model arguments. The three kinds of premises listed above are meant to model the different kinds of critical questions that might be asked in expert judgment contexts for use in an argument scheme. Overall, this article provides a detailed account of how some parts of the expert judgment process may be supported via rigorous argument.

42. Walton, D.N. 2006. Examination dialogue: A framework for critically questioning an expert opinion. *Journal of Pragmatics* 38: 745–777.

Keyword(s): argumentation, article, formal methods

This article makes an effort of building upon work in the field of argumentation theory that has used types of dialogue as contexts of argument use. It analyzes a type of dialogue known as examination dialogue, “in which one party questions another party, sometimes critically or even antagonistically, to try to find out what that part knows about something” (745). This sort of dialogue can be used to critically question arguments based on expert opinion. The examples studied in the article are: “(1) exegetical analyses and criticisms of religious and philosophical texts, and (2) legal examinations and cross-examinations conducted in a trial setting” (745).

The article’s concluding section discusses the results of the preceding analysis. It asserts that examination dialogues have two general goals: “extraction of information” and “testing of the reliability of the information extracted from the respondent” (772). Difficulties regarding the latter goal are discussed: “This process is non-trivial, because experts often use technical language, and what they say is often reported second hand, and not quoted directly” (773). The prevalence of examination dialogues in other contexts such as artificial intelligence and law is mentioned, and it is suggested that analysis of examination dialogues in the critical questioning of arguments from expert opinion might apply to those contexts as well and prove useful in various practical applications.

43. Walton, D.N. 1997. *Appeal to expert opinion: Arguments from authority*. University Park, PA: Penn State University Press.

Keyword(s): argumentation, book, formal methods

This is a textbook. No annotation is provided.

44. Wittmann, M. E., Cooke, R. M., Rothlisberger, J. D., Rutherford, E. S., Zhang, H., Mason, D. M., & Lodge, D. M. (2015). Use of structured expert judgment to forecast invasions by

bighead and silver carp in Lake Erie. *Conservation Biology*, 29(1), 187–197.  
<http://doi.org/10.1111/cobi.12369>

Keyword(s): article, elicitation, formation, use

This article deals with the use of expert judgment in a fairly specific domain. It is noted that the relevant context of judgment is one in which “the data or resources necessary... are incomplete or absent” (187). The structured expert judgment process used involves numerous experts whose judgments are aggregated. A notable aspect of this aggregation is the differential weighing of experts based on the statistical accuracy of their judgments and their “informativeness through performance measurement on a set of calibration variables” (187). The general conclusion of the paper is that “SEJ [(Structured Expert Judgment)] can be used to quantify key uncertainties of invasion biology and also provide a decision-support tool when the necessary information for natural resource management and policy is not available” (187).